Quantitative Trait Loci Mapping of Seed Hardness in Soybean

Bo Zhang, Pengyin Chen,* Charles Y. Chen, Dechun Wang, Ainong Shi, Anfu Hou, and Tetsuaki Ishibashi

Abstract

Soybean [*Glycine max* (L.) Merr.] seeds with undesirable texture cause processing complications in soyfood production. Seed hardness is an important quality attribute for food-grade soybeans. The objective of this study was to identify quantitative trait loci (QTLs) associated with seed hardness in soybean. Three generations of F_2–derived lines (159 $F_{2:3}$, $F_{2:4}$, and $F_{2:5}$ lines) from a soft ('SS-516') × hard ('Camp') soybean cross were grown in a replicated test in Fayetteville, AR, in 2003, 2004, and 2005. Pressure-cooked samples from each line were tested for seed hardness using a texture analyzer. A total of 874 simple sequence repeat (SSR) markers were used to screen the parents; 177 out of 236 polymorphic markers between the parents showed polymorphism in the $F_{2:3}$ lines. A linkage map for seed hardness was established using 148 SSR markers, 15 of which were new and added to the current public soybean genetic linkage map. All identified markers were placed on 19 linkage groups (LGs) and covered 1363.7 cM of the soybean genome with an average distance of 9.6 cM between markers. Broad-sense heritability was estimated to be 0.56 for seed hardness. Two stable QTLs across environments (*Ha1* and *Ha2*, $p < 0.00001$) were identified near Satt229 on LG L and Satt531 on LG D1a, respectively, for the average seed hardness over 3 yr. *Ha1* had a logarithmic odds score of 6.17 with $R^2 = 12.7\%$; *Ha2* had a score of 5.08 with $R^2 = 36.1\%$. A dominance-by-dominance interaction was detected between *Ha1* and *Ha2*, explaining 7.9% of the phenotypic variance. Research is under way to confirm the identified QTLs for soybean seed hardness in multiple populations with different genetic backgrounds.

B. Zhang, P. Chen, A. Shi, A. Hou, and T. Ishibashi, Dep. of Crop, Soil, and Environmental Sciences, 115 Plant Science Building, Univ. of Arkansas, Fayetteville, AR 72701; C.Y. Chen, USDA-ARS, National Peanut Research Lab., Dawson, GA 39842; D. Wang, Dep. of Crop and Soil Sciences, A384-E Plant and Soil Science Building, Michigan State Univ., East Lansing, MI 48824. Received 2 Oct. 2007. *Corresponding author (pchen@uark.edu).

Abbreviations: CIM, composite-interval mapping; MAS, marker-assisted selection; MIM, multiple-interval mapping; QTL, quantitative trait locus; RFLP, restriction fragment length polymorphism; SMA, single-marker analysis; SSR, simple sequence repeat.

Soybean [*Glycine max* (L.) Merr.] not only is an important protein source for livestock and poultry but is also used for edible vegetable and industrial oil in the United States. (Rao et al., 2002). The demand for various soyfoods has been increasing due to their nutriceutical and phamaceutical implications such as reducing blood serum cholesterol and the risk of cardiovascular diseases (Physicians Laboratory, 2006). Soyfood has been consumed in Asia for over 1000 yr and is becoming increasingly popular around the globe. Therefore, breeding specialty soybeans for the soyfood market may be pivotal to the global soybean industry.

Food-grade soybeans are grouped into two major classes on the basis of soyfood manufacture. One is for whole-soy foods including natto, vegetable soybeans, and soy nuts; the other is for cracked-soy foods such as tofu, miso (fermented soup-base paste), soymilk, and soy sauce. Natto is a typical and popular Japanese fermented soyfood made of small seeds with soft and sticky texture, and it has higher isoflavone (genistein and genistin) content than soymilk or tofu (Fukutake et al., 1996). Vegetable soybean,

Published in Crop Sci. 48:1341–1349 (2008).
doi: 10.2135/cropsci2007.10.0544

677 S. Segoe Rd., Madison, WI 53711 USA

also called *edamame* in Japan and *mao dou* in China, is a green large-seeded soybean with a sweet, nutty flavor, and usually harvested at the R6 stage (Fehr et al., 1971). The important quality properties of vegetable soybean are seed color, texture, and size (Mbuvi and Litchfield, 1995). Soy nuts are roasted soybean seeds requiring soft texture after roasting and a low percentage of antinutritional factors such as trypsin inhibitors (Chauhan et al., 2003). Apparently, seed texture is an important factor in sensory evaluation of seed characteristics for whole-soy foods consumption. Seed texture is also important for cracked-soy foods because the texture of raw seeds and soyfood products will have an impact on the manufacturing process and consumer acceptance (Mullin and Xu, 2001).

Seed hardness is an important quality attribute for soyfoods because it affects water absorption, seed coat permeability, cookability, and overall texture. For example, seed hardness was found to be significantly correlated to steamed seed hardness in a natto quality-trait study (r = 0.75; Chandler et al., 2000). However, testing for seed hardness is costly and time-consuming and often not practical for breeders' selection at the early stages of a breeding program. The suitability of potential cultivars or seed product for the natto market is usually determined by professional testers based on sensory evaluations. At least 0.91 kg of soybean seeds is required for each sensory test in which 10 well-trained researchers taste natto products in a 4-wk period with four replications of each product (Wei and Chang, 2004). In addition, there is no standard method in testing soybean seed hardness using scientific instruments. A puncture test is one of the simplest and commonly used instrumental measurements for food texture (Bourne, 2002), but it can give rise to large experimental errors if only a single sample is tested. The shear test also has limited utility without an accepted standard testing procedure. A compression test is rarely used because gases are easily trapped in most products (Wheeler et al., 1994).

Several hardness testing methods were compared using a 16-blade shear cell, a single blade, a 2-mm probe, a 75-mm cylinder, and pea rigs (16 2-mm probes) with different amounts of seed (Zhang et al., 2008). All these methods require a texture analyzer and are not feasible for breeders to use in selection because measuring seed hardness is difficult and time-consuming.

Although seed hardness is one of the major traits evaluated for food-grade soybeans, little information is available on genetic analysis of seed hardness. Keim et al. (1990) identified five restriction fragment length polymorphism (RFLP) markers associated with seed germination affected by hard seeds in a population derived from a cross between *G. max* and *G. soja*. These markers accounted for a total of 71% of variation in hard seededness, which affects seed germination. Keim et al. (1990) also stated that three interactions between five QTLs represented 38% of the total hard seededness variation.

In soybean, DNA markers have been widely and frequently used in identification of QTLs for major seed quality traits such as protein, oil, and seed size. However, research on soybean seed hardness is lacking. Information on genetic control of seed hardness will help understand this important seed quality trait and accelerate the process of breeding specialty soybeans for the soyfood market. The objective of this study was to identify QTLs associated with seed hardness, which would improve breeding efficiency with marker-assisted selection (MAS).

MATERIALS AND METHODS

Plant Material

A total of 159 lines representing three generations of F_2–derived lines ($F_{2:3}$, $F_{2:4}$, $F_{2:5}$) were developed from a cross of 'SS-516' × 'Camp'. Both parents were small-seeded cultivars specifically developed for the natto market. Camp, released by Virginia Polytechnic Institute and State University, is hard, with a hardness of 580 to 640 N per 30 g cooked seeds (maximum force in newtons to compress 30 g cooked seeds). SS-516, released by the Southern States seed company, is soft, with a hardness of 340 to 400 N per 30 g cooked seeds. SS-516 has white flower and broad leaf, whereas Camp has purple flower and narrow leaf. In crossing, a single plant of SS-516 was used as female to mate with a single plant of the pollen parent Camp in a greenhouse in spring of 2002. Four F_1 plants were grown in the field and identified as true hybrids in summer of 2002. The four hybrid plants were harvested in bulk to form the F_2 population. The F_2 plants were then grown in the field and harvested individually to derive the $F_{2:3}$ lines in a winter nursery in Costa Rica. Subsequently, each line was maintained as a separate entry using bulked seed for phenotyping at the F_4 and F_5 generations.

All the field experiments were conducted at the University of Arkansas Agricultural Experiment Station, Fayetteville, AR. In summer 2003, 2004, and 2005, $F_{2:3}$, $F_{2:4}$, $F_{2:5}$ lines were planted in the field with a randomized complete block design with two replications. Each line was grown in a single-row plot 1.5 m in length, with a 0.95-m row spacing. Both Camp and SS-516 were included as checks in the test. A bulk leaf sample consisting of a young leaf per plant from each $F_{2:3}$ line was collected at the V5 growth stage (Fehr et al., 1971) and stored at −80°C for DNA extraction. Plants in each plot were harvested in bulk, and a 50-g sample was used for hardness analysis.

Seed Hardness Testing

Fifty grams of unbroken seeds with uniform seed size from each F_2–derived line was weighed and soaked in heat-resistant plastic boxes containing 250 mL water at ambient temperature for 16 h. Seeds were recovered from the soaking water with a sieve and blot-dried on paper towels. Stone seeds, which did not absorb water during soaking, were picked from the soaked seeds. Soaked-seed samples were pressure-cooked at 121.1°C and 1.2 kg cm^{-2} for 20 min. Hardness tests of 30 g cooked seeds from each entry were conducted in two replications using a

TMS Texture System (TMS-2000, Food Technology Corp., Sterling, VA) equipped with a 16-blade shear cell. The maximum force to compress cooked seeds in newtons was determined as seed hardness (Song et al., 2003).

Molecular Marker Screening

Total genomic DNA was isolated from ground frozen leaf tissue using the CTAB (hexadecyltrimethyl ammonium bromide) method (Kisha et al., 1997). Polymerase chain reactions (PCRs) were performed as described by Cornelious et al. (2005). The amplified product was loaded in a 6% (w/v) nondenaturing polyacrylamide gel and stained with ethidium bromide (Wang et al., 2003). Two parents were screened with 874 simple sequence repeat (SSR) markers randomly distributed on 20 linkage groups (LGs). Two hundred and thirty-six polymorphic markers between parents were used to genotype the $F_{2:3}$ lines.

Data Analysis

Phenotypic data on seed hardness was analyzed using JMP 5.0 software (SAS Institute, 2002) to remove outliers (studentized residual > 3 or < −3). Shapiro-Wilks' (W) test was used to test the mapping population for normal distribution. ANOVA was used to assess the difference of hardness among lines over 3 yr (2003 to 2005). Heritability of seed hardness was estimated using $H^2 = \sigma^2_g/[\sigma^2_{rg} + (\sigma^2_{ge}/e) + (\sigma^2/re)]$ (Nyquist, 1991), where H^2 is heritability, σ^2_g is genotypic variance, σ^2_{ge}/e is genotype × environment interaction variance, σ^2 is error variance, r is number of replications, and e is number of environments, which is the number of years in this study.

The SSR marker segregation data of $F_{2:3}$ lines were used to construct a genetic map by using JoinMap 3.0 (Van Ooijen and Voorrips, 2001) with the Kosambi function (Kosambi, 1944). Markers that exhibited unusual segregation ratios in the mapping population were excluded from the linkage analysis. Linkage parameter was set as recombination fractions of 0.40, likelihood (logarithmic odds [LOD]) score of 3.0, and the error detection ratio of 5%. QTL Cartographer (Basten et al., 1999) with single-marker analysis (SMA), composite-interval mapping (CIM), and multiple-interval mapping (MIM) analysis were used to identify QTLs associated with seed hardness by comparing genotyping data of $F_{2:3}$ lines and phenotyping data of three sets of F_2–derived lines ($F_{2:3}$, $F_{2:4}$, $F_{2:5}$). In SMA, $p < 0.001$ was used as the threshold for significant markers. In the CIM analysis, the empirical significance threshold was determined by 1000-time permutation with a walk speed of 1 cM and significance level of 0.05 (Churchill and Doerge, 1994). Multiple-interval mapping analysis was used to estimate the optimum positions and effects of QTLs as well as QTL epitasis interactions. The MIM model $c(n) = \ln(n)$ was selected with a walk speed of 1 cM. Mapchart (Voorrips, 2002) was conducted to create the LOD plots according to the data from QTL Cartographer.

RESULTS

Seed Hardness

The seed hardness of 159 F_2–derived lines in each year from 2003 to 2005 and 3-yr average showed continuous variation and normal distribution, suggesting that seed hardness is a quantitative trait controlled by multiple genes (Fig. 1). ANOVA results showed that the seed hardness between two of three years and the interaction of genotype and year were significantly different (data is not shown). Hardness data in each year and combined in 3 yr were analyzed. The seed hardness of $F_{2:3}$ lines ranged

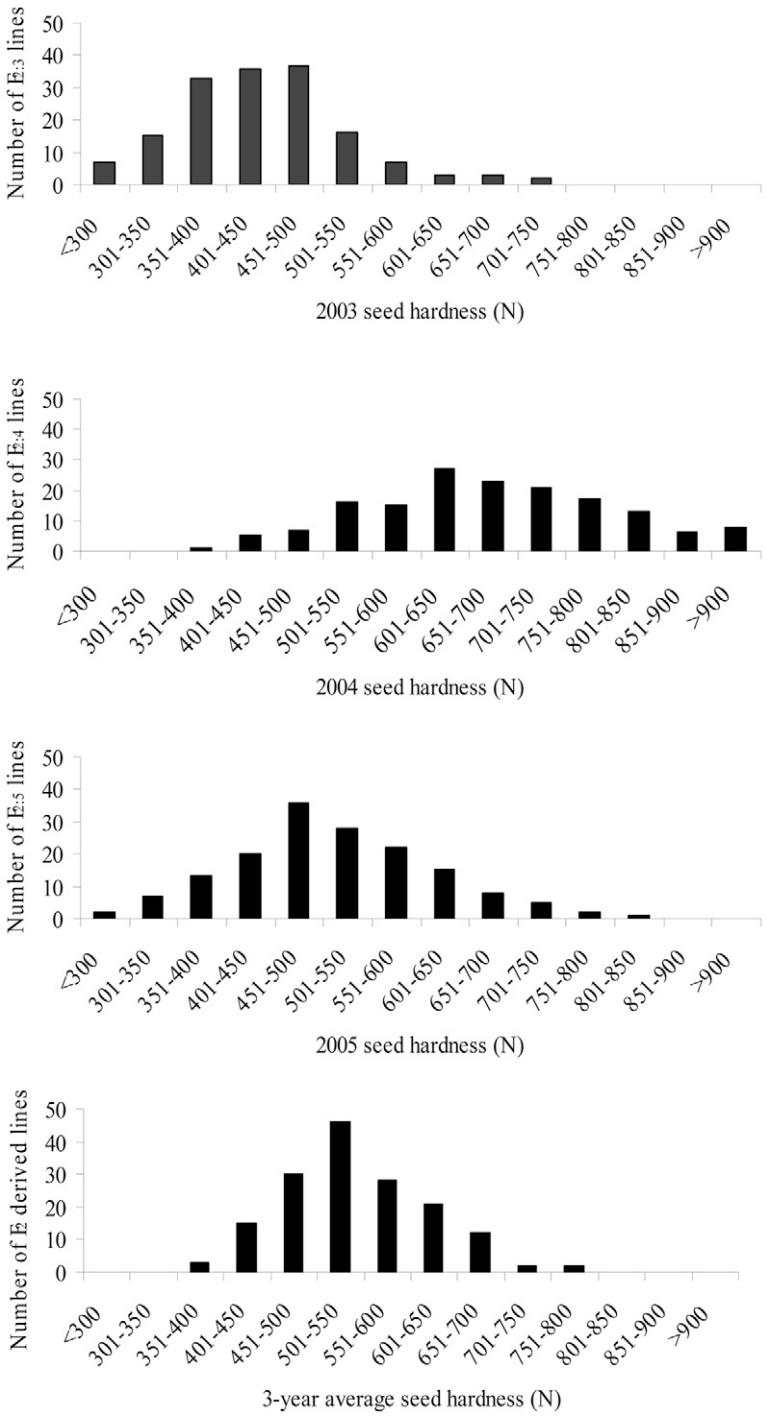

Figure 1. Distribution of seed hardness of $F_{2:3}$ lines in 2003, $F_{2:4}$ lines in 2004, and $F_{2:5}$ in 2005, and average seed hardness over 3 yr.

from 246.5 to 744.5 N with a mean of 438.8 N, the $F_{2:4}$ lines ranged from 393.5 to 972 N with a mean of 677.7 N, and the $F_{2:5}$ lines ranged from 297.5 to 843 N with a mean of 544.7 N. The 3-yr average of the F_2–derived lines ranged from 368.5 to 799.5 N with a mean of 544.7 N. The broad-sense heritability (H^2) estimate using variance components for the seed hardness in 3-yr average data of the mapping population was 0.56.

Linkage Map Construction

Of the 874 SSR markers screened, 236 markers showed polymorphism between the parents, Camp and SS-516. One hundred and seventy-seven polymorphic markers (75%) that were easy to score and had good amplification segregated in the mapping population, of which 148 markers were integrated into all but LG D1b of the 20 LGs with a total genomic coverage of 1363.7 cM (Table 1). The grouping and position of most markers were consistent with the public soybean linkage map (Song et al., 2004). Each LG contained an average of eight markers with an average interval of 9.6 cM. Linkage groups G, L, and M had the best coverage with more markers for their relatively long genetic distance. There were 18 markers on LG G, for 101.5-cM genomic coverage, and LGs L and M included 16 and 15 markers for 80.4 and 114-cM genomic coverage, respectively. Fifteen new markers that are currently not on the published public consensus map were mapped on the linkage map of this study (Table 1). Specifically, five markers (Satt704, Satt098, Satt401, Sat_210, and Satt616) were added on LG G, four (Sct_029, Satt176, Satt040, and Satt093) on LG F, two (Satt714 and Satt265) on LG L, and one on each of LGs B1 (Sat_257), C1 (Sat_327), H (Satt246), and N (Satt084) (Table 1).

Quantitative Trait Loci Analysis

Quantitative trait loci for seed hardness were analyzed separately in each year and jointly across 3 yr. In the SMA, 34 markers in 2003, 5 markers in 2004, and 68 markers in 2005 were significantly associated with seed hardness ($p < 0.001$). Twenty-nine markers were shared between 2003 and 2005, two markers were shared between 2003 and 2004, and three markers were shared between 2004 and 2005 (Table 1). Sat_110 and Satt531 on LG D1a were shared among 2003, 2004, and 2005 (Table 2).

In the CIM and MIM analysis, empirical significance threshold was computed as LOD score of 5.26 after 1000 permutations in the mapping population in 2003. However, none of markers that were significantly associated with seed hardness in 2003 had an LOD score as high as or higher than 5.26. Satt229 (72.3 cM) on LG L had the highest LOD score of 4.47 among all the markers, and it was significantly associated with seed hardness in 2003 ($p < 0.001$) in SMA (Fig. 2a). In 2004, the LOD score of 3.73 was used as the empirical significance threshold. A putative QTL (LOD = 4.5) was located at 120.4 cM on LG D1a, which is close to Satt531 at 112.4 cM (Fig. 2b). This QTL accounted for 24.4% of the seed-hardness variation. In 2005, the same two QTLs were identified to be significantly associated with seed hardness in the mapping population (Fig. 2c and d). One QTL resides at 72.3 cM on LG L where Satt229 is exactly located ($p < 0.00001$ in SMA) in our genetic linkage map (Fig. 2c). The other QTL was located at 110.3 cM on LG D1a between Satt532 (95.3 cM) and Satt531 (112.4 cM; $p < 0.000001$ in SMA) (Fig. 2d). Thus, Satt531 is likely the marker tightly linked to this QTL on LG D1a. These two QTLs accounted for 9.4 and 8.2% of phenotypic variation, respectively, which is lower than the variation explained by Satt229 (10.7%) in 2003 and by Satt531 (24.4%) in 2004. The slight discrepancy in QTL effect is probably because these two QTLs exhibited epistatic interaction on seed hardness in this population, and this interaction occurred under specific environments.

In summary, Satt229 was marginally associated with the QTLs for seeds hardness on LG L, with a slightly lower LOD score than the threshold in 2003. Satt229 was not significantly associated with seed-hardness variation in 2004 (LOD = 1.37) but was tightly linked to the seed hardness QTLs on LG L in the mapping population in 2005. Satt531 was significantly associated with seed-hardness variation but had a low LOD score (3.38) in 2003. However, Satt532 was tightly linked to the QTLs on LG D1a in 2004 and 2005.

Two QTLs were identified near Satt229 on LG L and Satt531 on LG D1a, respectively, for the average seed-hardness over 3 yr (Fig. 3a and b). The first QTL, designated *Ha1,* was located at 72.3 cM on LG L and 0.0 cM from Satt229. *Ha1* had an LOD score of 6.17, which was higher than the empirical significance threshold of 4.13 with R^2 = 12.7%. The other QTL, named *Ha2,* was located 136.4 cM on LG D1a, 24 cM from Satt531 and 18.5 cM from Satt184. *Ha2* had an LOD score of 5.08 with R^2 = 36.1%.

MIM models further confirmed the two QTLs for seed hardness, defined their position on the linkage map, and revealed their epistatic interaction effect on phenotypic variation. *Ha1* was still mapped at 72.3 cM on LG L, while *Ha2* was relocated at 128.4 cM on LG D1a, 16 cM away from Satt531. A dominance-by-dominance interaction was detected between *Ha1* and *Ha2*, and it explained 7.9% of the phenotypic variance for seed hardness.

DISCUSSION

Soyfood manufacturers prefer soft small-seeded soybeans for natto production. Breeding populations such as the mapping population used in this study will enable breeders to select superior natto lines with desired seed texture. In this study, there were significant differences in seed hardness among lines in the population in 3 yr, with significant interaction of year by genotype. The mean seed hardness in 2004 was higher than that in both 2003 and

Table 1. Soybean genetic map of seed hardness including 148 simple sequence repeat (SSR) markers with estimated position on respective linkage groups (LG).

Locus	Year†	LG	Position	Locus	Year	LG	Position	Locus	Year	LG	Position
			cM				cM				cM
Satt572		A1	0	Satt256	5	D2	72.7	Satt596		J	41.6
Satt449		A1	3.4	Satt514	3	D2	85.5	Scaa003		J	48.7
Satt454	5	A1	18.8	Sat_354		D2	90.2	Satt456		J	55.1
Sat_356	5	A1	36.2	Satt369	3, 5	E	0	Sat_395	3, 5	J	90.9
Satt050	4, 5	A1	44.9	Sat_380	5	E	15.7	Satt475		K	0
Sat_407	5	A1	53.7	Satt699	4	E	23.5	Satt102		K	42.4
Sat382		A2	0	Satt212	5	E	49.6	Sat_408		L	0
Satt429	5	A2	3.1	Sct_029‡		F	0	Satt523		L	17.1
Satt228	5	A2	14.4	Satt176‡	3, 5	F	9.6	Sat_405		L	21.4
Sat_377	3, 5	A2	21.8	Satt343	3, 5	F	19.1	Sat_388	5	L	22.7
Sat_392	3, 5	A2	34.1	Satt569	5	F	24.4	Satt418		L	24.1
Satt089	3, 5	A2	49.6	Satt040‡	3, 5	F	31.9	Sat_397		L	25.6
Satt133		A2	57.2	Satt252	3, 5	F	43.9	Satt313		L	27.8
Satt341		A2	62.1	Satt335		F	47.8	Satt714‡		L	31.2
Satt437		A2	88.8	Satt093‡	3, 5	F	57.7	Satt265‡		L	35.7
Sat_257‡	3, 5	B1	0	Sct_188		F	62.9	Satt462	5	L	49.3
Satt359	3, 5	B1	6.4	Sct_033	5	F	83.4	Satt156	5	L	55.4
Satt597		B1	17.7	Sat_117		G	0	Sct_010	3, 5	L	61
Sat_095		B1	27.2	Satt704‡	5	G	10.4	Sat_113	5	L	67.6
Sat_360		B1	41	Satt503		G	13	Satt229	3, 5	L	72.3
Satt197		B1	46.5	Satt504	3	G	20.2	Satt513		L	76.9
Satt304		B2	0	Satt564	3	G	26.1	Satt373	3	L	80.4
Sat_355		B2	11.5	Sat_403		G	39.2	Satt551		M	0
Satt070		B2	20.8	Satt324		G	46.2	Satt680		M	28.2
Satt556		B2	44.7	Satt098‡		G	50.3	Satt346		M	30.5
Satt195		C1	0	Satt401‡		G	55.2	Satt336		M	38.6
Satt399		C1	7.1	Sat_168	3, 5	G	60.1	Satt220		M	41.6
Sat_042	5	C1	31.1	Satt594		G	62.9	Sat_288		M	51.4
Satt294	3, 5	C1	35.3	Satt570		G	63.2	Satt175	5	M	57.7
Satt361	5	C1	42.3	Satt224‡	3, 5	G	67.5	Satt536	5	M	63
Satt190	5	C1	44.3	Sat_210	3, 5	G	69.8	Satt150	3, 5	M	66
Sat_327‡		C1	54.9	Satt616‡		G	75.7	Satt245	5	M	69.3
Satt202		C2	0	Satt217	5	G	82	Satt463	5	M	71.4
Satt286		C2	27.5	Satt038		G	97.9	Satt567		M	83.1
Satt071		D1a	0	Satt610	3, 5	G	101.5	Satt677	3, 5	M	90.7
Satt147		D1a	26.5	Satt293		H	0	Sat_391		M	104.3
Sat_345	4	D1a	45.2	Sat_205	3, 5	H	8.6	Sat_389		M	114
Sat_110	3, 4, 5	D1a	64.1	Satt246‡		H	43.3	Satt084‡		N	0
Satt267		D1a	76	Sctt_009	5	H	52.5	Satt125		N	29.2
Satt254		D1a	84.5	Sat_218	5	H	57.9	Satt080		N	40.1
Satt532		D1a	95.3	Satt353	5	H	65.1	Satt387	5	N	48
Satt531	3, 4, 5	D1a	112.4	Satt434	5	H	79	Satt521	5	N	74.6
Satt184		D1a	154.9	Satt496		I	0	Satt257	5	N	109.1
Sctt_008	3	D2	0	sat_174		I	3.1	Sat_108	5	O	0
Sat_022	5	D2	29.1	Satt354	3, 5	I	17	Satt243	5	O	22.8
Satt461	5	D2	48.6	sat_170	3, 5	I	36.7	Satt581		O	29.2
Satt311		D2	52.3	Satt330	3, 5	I	38.6	Satt331		O	37.9
Satt528	3, 5	D2	54.7	Satt292	3, 5	I	40	Satt123		O	49.7
Satt082	5	D2	58.7	Sat_228	5	J	0	Satt477		O	58.8

†Numbers indicate presence of SSR marker–quantitative trait locus association in the corresponding population in specific years. 3, 2003; 4, 2004; and 5, 2005.

‡New SSR markers identified in this study but not mapped on the published consensus map.

Table 2. Shared single sequence repeat markers significantly associated with seed hardness in F_2–derived lines ($F_{2:3}$, $F_{2:4}$, and $F_{2:5}$) from soybean cross SS-516 × Camp from 2003 to 2005.

Marker	LG†	Position		LOD§		
		In present study	On consensus map‡	2003	2004	2005
		—— cM ——				
Sat_110	D1a	64.1	62.5	11.3	4.2	11.3
Satt531	D1a	112.4	40.9	8.6	5.2	4.2

†LG, linkage group.

‡The consensus map is the map by Song et al. (2004).

§LOD, base 10 log of likelihood ratio (logarithmic odds score).

2005, perhaps because there was less rainfall in August and September of 2004 (2.97 cm) at the R3–R7 stages of soybean development as compared with 2003 (16.81 cm) and 2005 (13.31 cm). Lack of rainfall was most likely the reason for harder seeds in 2004 than 2003 and 2005. The broad-sense heritability of seed hardness (0.56) on entry-mean basis was moderately high, similar to the heritability of other seed quality traits such as aspartic acid (0.57, Panthee et al., 2006) and oil (Brummer et al., 1997), and lower than that of protein (Brummer et al., 1997). Selection for proper seed texture can be achieved through genetic manipulation and breeding selection, although environmental conditions also play an important role in seed hardness. Since there is very limited information on genetic analysis of seed hardness in soybean, results from this study provide a fundamental basis for future research on genetic improvement of seed hardness of soybean.

Of the 874 SSR markers tested, 235 (26.9%) were polymorphic between the two parents, and 177 (20.2%) markers showed polymorphism in the $F_{2:3}$ population. This polymorphic ratio was in agreement with 16 to 30% polymorphism reported for other soybean populations (Arahana et al., 2001). In this study, the genetic linkage map covered 1363.7 cM of the soybean genome, which was equivalent to 54.0% genetic distance of the public linkage map (Song et al., 2004). Furthermore, 15 newly identified markers in the present study provided reference of markers added to the public linkage map, which improved the likelihood of identifying QTLs associated with seed hardness and would facilitate QTL discovery for other traits in the future. In the consensus map, Satt184 (17.5 cM) and Satt531 (40.9 cM) were 23.4 cM apart on LG D1a. Sat_353 (36.2 cM) was the only SSR marker located between Satt531 and Satt184 (Song et al., 2004). However, Sat_353 did not show polymorphism among lines in the mapping population of this study. Additional markers in the region between Satt531 and Satt184 will be needed to reduce the interval of the *Ha2* region in the future.

Useful QTLs in MAS should be stable across environments. Evaluating a QTL across several years can be used to assess stability (Brummer et al., 1997). The two QTLs identified in this study were detected using seed hardness data from 3 yr. *Ha2* was more stable than *Ha1* because *Ha2* was detected in both 2004 and 2005, while *Ha1* was detected in 2005 and was only a candidate QTL in 2003. However, both QTLs were detected if pooled data were used over the 3 yr. In addition, one of the seed coat hardness QTLs in Soybase, Hrd Sd 1–3, is in close proximity to *Ha1*. Therefore, both QTLs should be useful, once confirmed, in a marker-assisted breeding program. Similarly, QTLs for protein and oil were detected in five populations in 3 yr and in three populations in 2 yr (Brummer et al., 1997). Breeders may combine the two QTLs in MAS in early generations to improve the selection efficiency and accelerate breeding process for seed hardness under different environments.

Quantitative trait locus regions usually contain loci for more than one trait (Wang et al., 2004). Satt229, linked to *Ha1* on LG L, also flanked a QTL for oil content (Hyten et al., 2004). The other flanking marker of this oil QTL was Satt373, which is 8.1 cM from Satt229 in our linkage map. Several seed-size QTLs were detected upstream of Satt229 (93.9 cM) in the consensus map (Hoeck et al., 2003). Satt006 (92.0 cM) and Sat_099 (78.2 cM) on LG L explained 32.5% and 33.7% of seed size variation (Hoeck et al., 2003). Quantitative trait locus oil-5 was located between Satt184 and Satt179 on LG D1a (Hyten et al., 2004). Satt531 (40.9 cM) was in the middle of Satt184 (17.5 cM) and Satt179 (56.2 cM) on LG D1a in the consensus map. Thus, *Ha2,* tightly linked to Satt531 in our study, was located within this region. Panthee et al. (2006) also identified a seed-size QTL near Satt184 with R^2 of 11.3, and a lysine QTL near Satt184 with R^2 of 13.1. These regions of LGs L and D1a, with multiple QTL clusters for seed-quality traits, may be attributed to genetic linkage and/or pleiotropy. Nevertheless, it is possible that some of the molecular markers identified in these regions can be used in MAS for multiple traits.

A dominance by dominance epistatic interaction between *Ha1* and *Ha2* existed in controlling seed hardness and explained 7.9% of the phenotypic variation, which is useful in breeding for proper seed hardness. Both QTLs may be needed in a correct allelic form to give rise to maximum genetic effect on seed hardness in a given line. Similarly, three interactions between five QTLs were identified to represent 38% of the total hard seededness variation in a previous study (Keim et al., 1990). The gene controlling hard seededness in soybean at the i locus (between pT-153a and pA-111 markers) was thought to epistatically manipulate other genes (Bernard and Weiss, 1973). Thus, the interaction between pA-111 and the other three markers confirmed the regulatory function of the i locus. Continued research efforts are needed to validate the two QTLs for seed hardness indentified in this study

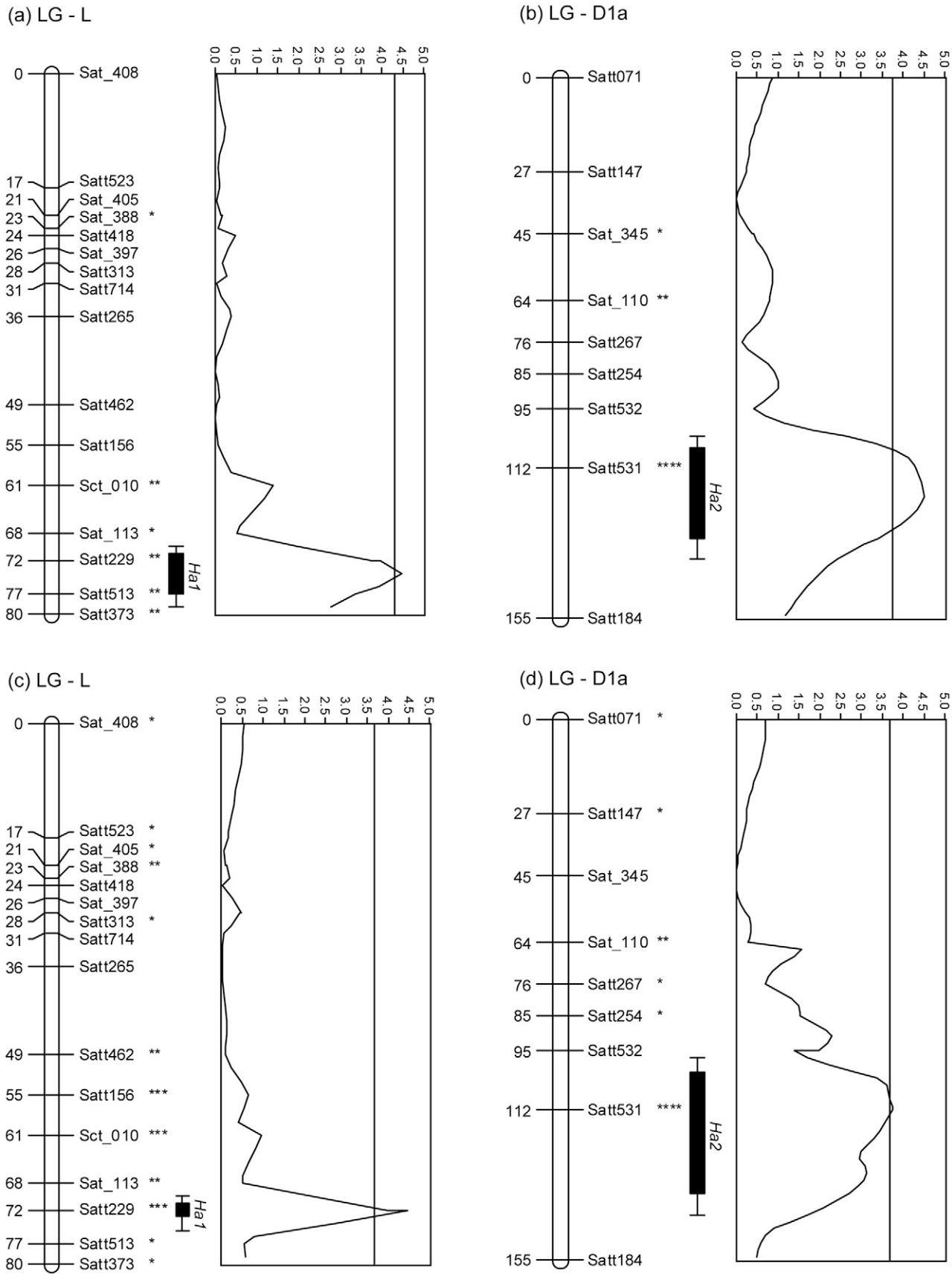

Figure 2. Logarithmic odds score plots for seed-hardness QTLs on linkage groups (LGs) L and D1a. Markers with *, **, ***, and **** are significantly associated with seed hardness in single-marker analysis at $p = 0.01$, 0.001, 0.0001, and 0.00001 levels, respectively. Linkage group L (a) shows a QTL for seed hardness in 2003; LG D1a (b) shows a QTL for seed hardness in 2004; LG L (c) and LG D1a (d) show QTLs for seed hardness in 2005.

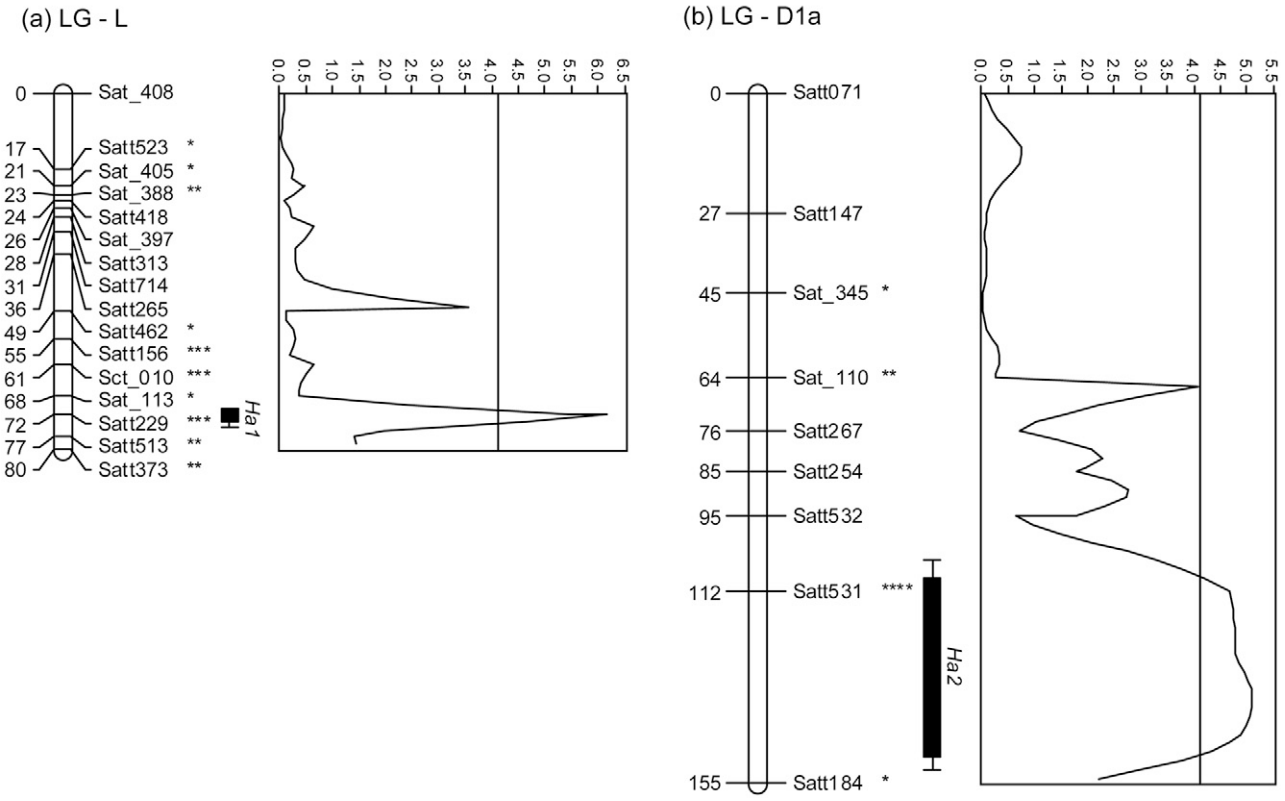

Figure 3. Logarithmic odds score plots for seed-hardness QTLs on linkage groups L (*Ha1*) and D1a (*Ha2*), using average seed hardness data over 3 yr. Markers with *, **, ***, and **** are significantly associated with seed hardness in single-marker analysis at the 0.01, 0.001, 0.0001, and 0.00001 levels, respectively.

so that they can be effectively used in MAS for seed hardness of specialty soybeans for natto production.

References

Arahana, V., G.L. Graef, J.E. Specht, R. Steadman, and K. Eskridge. 2001. Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. Crop Sci. 41:180–188.

Basten, C.J., B.S. Weir, and Z.B. Zeng. 1999. QTL cartographyer, version 1.13. Dep. of Statistics, North Carolina State Univ., Raleigh.

Bernard, R., and M.U. Weiss. 1973. Qualitative genetics. p. 112–154. *In* B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. ASA, Madison, WI.

Bourne, M.C. 2002. Food texture and viscosity: Concept and measurement. 2nd ed. Academic Press, San Diego, CA.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370–378.

Chandler, W.G., W.R. Ferh, and L.A. Wilson. 2000. Association of soybean seed traits with physical properties of natto. Crop Sci. 40:1529–1534.

Chauhan, O.P., G.S. Chauhan, A. Sharma, A. Bajpai, and D.N. Yadav. 2003. Effect of varieties on the quality characteristics of roasted soybean. Int. J. Food Prop. 6:281–289.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Cornelious, B., P. Chen, Y. Chen, D. Leon, J.G. Shannon, and D. Wang. 2005. Identification of QTLs underlying water-logging tolerance in soybean. Mol. Breed. 16:103–112.

Fehr, W.R., C. Caviness, D. Burmwood, and J. Pennington. 1971. Stage of development descriptions for soybeans. Crop Sci. 11:929–931.

Fukutake, M., M. Takahashi, K. Ishida, H. Kawamura, T. Sugimura, and K. Wakabayashi. 1996. Quantification of genistein and genistin in soybeans and soybean products. Food Chem. Toxicol. 34:457–461.

Hoeck, J., W.R. Fehr, R.C. Shoemaker, G. Welke, S. Johnson, and S. Cianzio. 2003. Molecular marker analysis of seed size in soybean. Crop Sci. 43:68–74.

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, and M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109:552–561.

Keim, P., B.W. Diers, and R.C. Shoemaker. 1990. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79:465–469.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. 12:172–175.

Mbuvi, S.W., and J.B. Litchfield. 1995. Green soybeans as a vegetable: Comparing green soybeans with green peas and lima beans; and maximized harvest time determinations using mathematical modeling. J. Veg. Crop Prod. 1:99–121.

Mullin, W.J., and W. Xu. 2001. Study of soybean seed coat components and their relationship to water absorption. J. Agric. Food Chem. 49:5331–5335.

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235–322.

Panthee, D.R., V.R. Pantalone, A.M. Saxton, D.R. West, and C.E. Sams. 2006. Genomic regions associated with amino acid composition in soybean. Mol. Breed. 17:79–89.

Physicians Laboratory. 2006. U.S. soyfoods directory: Soyfoods descriptions. Available at http://www.soyfoods.com/soyfoods-descriptions/descriptions.html; verified 22 Apr. 2008). Physicians Laboratory, Kernersville, NC.

Rao, M.S., A.S. Bhagsari, and A.T. Mohamed. 2002. Fresh green seed yield and seed nutritional traits of vegetable soybean genotypes. Crop Sci. 42:1950–1958.

SAS Institute. 2002. JMP statistics and graphics guide. Version 5.0. SAS Inst., Cary, NC.

Song, J., G. An, and C. Kim. 2003. Color, texture, nutrient contents, and sensory values of vegetable soybeans [*Glycine max* (L.) Merrill] as affected by blanching. Food Chem. 83:69–74.

Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, and P.B. Cregan. 2004. A new integrated genetic linkage map of soybean. Theor. Appl. Genet. 109:122–128.

Van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap version 3.0, software for the calculation of genetic linkage maps. Plant Research Int., Wageningen, the Netherlands.

Voorrips, R.E. 2002. Mapchart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78.

Wang, D., G.L. Graef, A.M. Procopiuk, and B.W. Diers. 2004. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor. Appl. Genet. 108:458–467.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828–1832.

Wei, Q., and S.K.C. Chang. 2004. Characteristics of fermented natto products as affected by soybean cultivars. J. Food Process. Preserv. 28:251–273.

Wheeler, T.L., M. Koohmaraie, L.V. Cundiff, and M.E. Dikeman. 1994. Effects of cooking and shearing methodology on variation in Warner-Bratzler shear force values in beef. J. Anim. Sci. 72:2325–2330.

Zhang, B., M. Tamura, J. Berger-Doyle, and P. Chen. 2008. Comparison of instrumental methods for measuring seed hardness of food-grade soybean. J. Texture Stud. 39:28–39.